California Outreach Visit – October 17th, 2012

Discussion on data sharing and the background and approach of WaDE

Value of exchanging data – There are many, many data pieces here in California. What's working and what's not working, what kinds of information do you need for your job? How important is sharing data to you and in an ideal world what would you like to see?

Rich – Publishing of future scenarios, every time they get a plan out, they get tons of data request, but the data sits in spreadsheets. Need to find a way to automate it and get it into a database, people need to be able to find it and understand the data.

Q - Would really avoid duplication of efforts on similar types of work, identifying the different kinds of effort (assessments) and minimize that duplication of effort.

Bruce – coordinating the data is of interest. WQX is a good example, a lot of exchanges, let's make it work together. We can emulate these larger exchanges perhaps.

Gary – Two drivers, staff are being asked to answer more complicated questions. New questions require more analyses and more data. Who is collecting what? How can you use it? This becomes very important. Initiative of the department is to promote integrated water management within the agency and within local districts, etc. First step to integrating is an assessment, formulate what kind of analyses are we already doing and what kind of analysis could we do in the future. Also, data collection is a large lumbering financial target. Budget cuts require cutting back on data collection. Data programs can't survive much longer in this scenario. To make it more efficient, you have to collaborate and share data. Maybe there's duplication, maybe there isn't, need to discover economies. Big props to Utah from Gary about data collection and lean system. Voluntary information is in a certain system, but we need a unified framework to make it easy. Water compliance is an issue, data exchange for water rights reporting. Need to share district generated data (have access to?).

Mary – state and federal provide funding that includes data collection, having a centralized system for people to share the data.

Loren – Standardization is imperative, 2010 estimated use reports trying to correlate from different sources is very difficult task for the GS.

Erin – Take a step back from how are we going to use the data and focus on using common nomenclature, what are the standard units, temporal, spatial, methods, documentation? Very interesting point about the idea that there is not a one-size-fits all. If you try to make your schema too strict you lose a lot of folks.

Mike – High data standards also have a cost, focus on ROI with regard to standard compliance. Need balance here between having great data and spending a reasonable amount of money to get it. Also, some systems are obsolete and need development, but often an agency only gets one shot to build the system and O&M (governance) is not considered. There is a new sunk cost for developing new standards or developing new infrastructure.

Q – Agree with the development of flexible framework of data standards and development. That seemed to help (more governance). "Let's learn to walk before we run."

Tim – ACWI.gov. Why re-invent the wheel? National Groundwater Monitoring Effort has an ongoing oversight on changes. May be a successful template to look at for WaterPIE.

Gary – In CA, there are 58 counties all in different states of development, need to try to set up a standard that they (and districts, etc.) and look to and build/project that into the future. Need a guideline standard. Also need to consider the sustainability of data infrastructure (server upgrades, bandwidth, etc.).

Kamyar – We should have a standard and plan for an evolution of participation, need something to serve as a translator between DWR's standard and the providers'. It is difficult to ask end-users to change everything, but we can create translators. Question about this. Translators can be good, but this can also impede change or upgrades, cause more problems. Need a way to do this without asking everybody to redo everything they've done in the past. NEIN (??), no more funding, something happened… should it move toward a WADE system?

Q - Adding different web service inputs requires more advanced metadata. Need for more documentation. Talk to the districts and send the survey to more of the CA people? CIMIS (?) for agriculture water use, do these use XML and web services?

Discussion of other data types for California and the need for a standard approach for the services – what does that URL look like? What are the parameters? Exchange network standardized the technology and the web services for the water quality exchange, but it depends on what type of information you're trying to retrieve. Some data are returned differently (?). That technology is out there for either a staging database set up or a new node or plug-in or for a virtual node. One issue is that the Exchange Network is missing the REST specification. Discussion of the REST evolution in the EPA.

Gary – WaterPIE – architecture management is a little chaotic. What if spokes are all over the place? Could incorporate too many. Maybe focus on a few of the essential spokes for now before adding in more diverse data inputs. Will these feed into the Water Census? Yes, hopefully, this gives the states a chance to gear up for that ahead of time. If the Western states can agree on the technology, then we can leverage the development and keep it moving forward. First need to review the different technologies used for data exchange, then parallel develop can then be utilized by whoever wants to use it.

Question about the staging database. These sit next to the state system. Web services code is built on top of the staging database. The real question is do you have someone on staff who can do either web services or transfer of data? This would dictate your approach. Staging helps with operation and won't interfere with the day to day operations of a database manager (network resource consumption). Question about long-term db population and maintenance. EPA has added new data flows and changes to schema each few years that are overseen with a governance process. Implement each dataflow as a 'plug-in.' WaDE uses the same approach, but on a much smaller and less formal scale. EPA has much more formal layers of governance to deal with proposed changes and maintenance.

Dwane talks about the EN web services 'get services' request and how it works. Big discussion on how to handle confidential data from the providers. Discussion about drawing the distinction between participating in WaDE that would likely not require authentication and WaterPIE. Kamyar mentions being a partner in the WaDE project. Tito? Is there a way for the user to drill down to a scale that compromises the privacy or the proprietary nature of the data? Within WaDE this will not be a problem,

but WaterPIE may need to consider this a constraint. Talked about the scales of CA water data: hydrologic analysis units, planning areas, watersheds, DAUs, then intersection with counties (!Very fine resolution for such a large state!)

…

10/18/2012 – Talk about WaterPIE

Tito explains the CA water portfolio, the "Beast." Based on region by region, broken down by modeling areas that coincide with HUC (3?). Cal water lines? Old area designation numbers that fit how the water moved. Works for all areas except for the valley. These don't correlate well, so switched to Detailed Analysis Unit (DAU). Set outlet of DAU at a point on the stream (nesting?). First bulletin done by DAU by county in 2009 report. Changed to an inflow-outflow style of accounting for water budgets. Mass balance or water budget approach. Top portion of the Beast is all the supply, break down into ag uses, M&I uses, and managed wetlands. Old days used to combine this. Now they break it out again. Get all the deliveries for that year. Break it down by project and categories. Estimate groundwater pumpage, some are measured, but some don't share the information. Sometimes they provide and sometimes not. Mapping lands that are served by groundwater, but only by one area (up north?). Trying to get to that point…. Values could be incorporated into the water budget and used to back-calculate the groundwater use out…. All the supply broken out by project/sector. And finally, they have estimated reuse from agriculture runoff. That gives total supply for the year. Consumptive use – "irrecoverable losses," but no one liked that term. Now just call it consumptive water. Have a section for deep percolation. It's like a checking account. When you add that with the supply and subtract uses you get the outflow. Track the outflow to the next unit. Does it become the inflow to the next DAU? Estimates for how those break down into counties. Usually a only small percentage off the streamgage down below. Don't have losses from tributaries, so that is estimated. This budget is done for each DAU/county. Check it, if there are errors, have to go back to the top. Aggregate the DAUs into counties for some customers (Loren?). Urban folks are breaking down the uses by a large degree, interior, exterior, large landscaping, single family, multi family, - called it 'synthetic data.' Three different types of aggregation at the state level, applied water use, net water use, and depletion, some interesting way of getting at groundwater uses… more discussion (need documentation for this). Depletion is consumptive use plus irrecoverable losses. Net water use – depletion = outflow. These values includes instream flow requirements and wild and scenic rivers, environmental data. Evelyn (Beast manager?) sums them up and adds them in. Information comes from past studies. Data program has been hit heavily. Departments have monthly deliveries of data. Water districts have their annual reports. Managed wetlands come from other agencies. Department of Health service requires reporting on the population, usage, surface, ground. A spreadsheet for each water purveyor. That's where you get the indoor/outdoor used. The analyses conducted are largely a matter of EXPERIENCE (local knowledge, local expertise – notion of why state data needs to be the source for studies and evaluations). May visit a different area on a cycle and totally refine the budget there. Example of a volcanic area where everything is deep perc., no runoff. A lot of teamwork, lots of folks checking and reviewing it for a gut check. Do have some monthly data, but put it out on an annual basis. Some of the results depend on the experience of the person doing the work at the local level. Is there a way to expedite the compilation of the data from the districts? Had a contract out that looked at all of the workflows for streamlining the process. Example – Ag water model that takes all different parameters and over years tries to predict the water balance. Information from that feeds into the ag portion of the 'Beast.' PWSS

(public water supply system) surveys. Don't map one to one. Forms ask for monthly, and aggregate to annually. When you send the surveys out, 800 different utilities, get 500 – 600 back from a variety, not always the same. Tito and his team end up teasing that information out and break it down. Fairly lengthy process, thought about making it more streamlined. Use judgment calls if something seems wrong, like a population not changing over a five year period. Know something's screwy. Also, have to deal with/account for un-organized areas with most people on self-supplied wells. Have to estimate that parameter. Over 7,000 – 8,000 water providers in the state. Often don't get any metadata about the surveys that get turned in. DWR is the agency adding any metadata. Have reports by hydrologic region, but might do it by planning areas (about 48). Water balance for the M&I, also do it for managed wetlands, but don't do it for the water purveyor – that may be our next phase would be to get the boundaries from purveyors. Data from purveyors are a more advanced report, which needs to be approved by their board of directors. 7-7 is encouraging purveyors to start looking at their water use field-by-field. Would it be helpful to do a balance for them? If purveyors/DWR did that, then they may only have to do it for the non-organized area. A lot of the ag data is not available (?), use Cal-ag model to estimate that. If they do start having more information, we have a system where this will plug in nicely. Q: not snowpack, but rainfall. Know that the snow is coming a little sooner the last 15 years. How long been doing the water balance budget? The last 10 years. The 2009 update has the figures from the last evaluation. Atmospheric river discussion – an event that dumps a lot of water that comes through Oregon, Washington, and northern California. Possibly think about how the data from the 50's correlates to the atmosphere river events as a back-check or calibration tool? Large-scale events, causes larger flooding. Only place that shows up in the budget is the Delta outflow. Precip? Yes, show up in the precip, but that feeds into the budget. The budget doesn't obviously show these events except for a large outflow out of the Bay. Just looking at the annual inflows and outflows. If you want to see the correlation between atmospheric rivers and the water supply check by reach on above normal/normal/below normal over the years.… River flow would show above normal flow across the board. Also, how full the river… break down by percentage – 80% full, 90% full, etc. WSWC likes to say that this is the western equivalent of the hurricane. The 'inland sea' the volume of the Mississippi River and that's why the whole valley flooded. Cal-ag model can be used weekly, monthly, yearly for consumptive use. Methods for Cal-ag, documented (need to get this). Basically the land-use program set up in Utah, survey the valley more often about 5 years or so, the rest of the state every 8 – 10 (?) years on rotation. Talked about viability of using remote sensing. Can find documentation of the process on the CA website. Cal-simita (?)… Get the flow diagram that will show how the data are created. Earlier, each region office had their own way of doing things – open up the black box and see how they were doing business. Question about how the Water Board determines whether water is 'available.' WB has a website with this type of information (need to review)(web services?). Up to an applicant to 'prove' that there's water available in that area. Needed to add environmental water to the water budget to try and add these values into the evaluation. DWR needs to make sure that their data aren't being used to suggest that there is water available from a legal perspective. Check to see whether the instream flows are being met. WB website… information can be used to show that you can use the water conjunctively. It'll be a while before the budget can include the groundwater component. Availability in the other states are adding together things like brackish groundwater, recycled water, reuse, saline, desalination, … would include these things and the recycled, reuse water as part of the supply… Karen Larsen has the water rights information. Have many of their data served as web services. E-permiting for water rights – real-time, monthly, weekly, information available through the board. Can set up a query for the DWR to get that information. Not required of all

submitters, but do have to do a water report. If they have time, they'll put that information into CELIX?? QA/QC maybe, but unknown.

BREAK

How can technology help us scope WATERPIE (WP)? We should talk about how the schema can be developed or updated so that we can cover the most overlap. How can we develop data that will be useful for WP. Dwane: One is how the WaDE schema is organized. Start with that. Ignore the more detailed portion of the schema (except some of the water use). Use a namespace that would contain all the WaterPIE idea. Organization is whoever owns the data. Report is a timeframe around which the data are organized, 2011, 2012. All information for that report. Reporting unit is your DAU or whatever it is that you're summarizing. Can use all five of the summary level data, plus your own namespace that includes whatever summary level data that they are interested in including. DWR would be the hub. Whoever is generating the data is the organization. Each organization is a spoke. Share widgets. Will the staging database idea work for WaterPIE?

Diagram on the board. SOAP vs. REST – need the expertise to support both architectures… so we should find something that works for our people or people we would hire. Plan would be to support both. Currently the EPA Exchange Network supports WSDL and SOAP, working on a REST standard. DWR has WSDL/SOAP, but doesn't support REST. In the coming year they will be supporting REST. With REST you can encrypt the channel, but would also have to encrypt the data. REST is good for Point to Point. Then it's okay to encrypt the channel. When you have more complex inquiries you need SSL – network would open and close the message. In order to avoid that, you encrypt the message itself. Considerations in choosing which technology you need. Can we use SOAP internally, but have a REST endpoint for any data put out to the public (i.e. WaDE)? This way we can make sure the data are secure that needs to be. Does the hub do messaging? What if networks drop? How to make sure the communication is occurring. Need to have messaging for the WaterPIE information. The overhead involved in authentication and tighter control goes up exponentially. Should use the business requirement to decide what WaterPIE's needs are. Need to look at the two different projects separately – WaDE is distinctly different from WaterPIE, but can look to each other for innovation and technology ideas. …

If the states can adopt a newer version of a schema or database, does that kill the central portal because other states 'break?' Depends on what the changes are like and how much they impact the states. Need to have governance for the WaterPiE schema also so that changes can be made. Within the EPA Exchange each dataflow will have an integrated project team, made up of people from the states. Think through how the changes are going to impact them. Changes get drafted and this goes through different governance levels. Communication group, technology review group, look at documentation and schemas, mid-level management group – look at resources and how it gets spent, upper-level management that helps allocate the funds for the smaller groups to make the changes. At WSWC, current council members will have the ENLC role, but we are missing the technical role.

BOARD Discussion

Discussion of common web services – need a flow configuration document. Discussion about how the exchange works, what the protocols are, what the input parameters are, how the services respond, how to

respond to errors. Once it's done, if you want to implement the data exchange, here's how that is done. Using our flow configuration document, you can frame yours based on that (depends on SOAP).

Discussion of how WaterPIE (with spokes) interfaces with WaDE. Discussion for different options via staging databases, web services or a web interface. Discussion of how SOAP works within the EN. SOAP and EN option: 1) free, standardized services, in place, can beg, borrow and steal 2) DWR may or may not already have the infrastructure in place to implement a SOAP-based enterprise repository, 3) Big question of whether your partners/data providers can deploy the node plugin, staging database, etc., WQX ended up giving a web-interface for people who could not do this, 4) QA/QC should be done at the level of the owner, there are reasons for the staging database as an additional QA/QC step, responsiveness of the partner doing it on-site as opposed to sending in copies of the DB or hosting a virtual node, (their refresh routine), 5) Concerns about changes through time from the partners/providers, need to archive the data as it comes in with some way of identifying the lineage of the data, 6) have a mandate to make the data available on the web, very big effect with respect to environmental compliance and with the water board. Once they're running, they need to plug into something, which should be first.

NEXT STEPS – for cooperating, Kamyar would like to plan on being a node on WaDE. DWR will start looking at WaDE schema. Have excel spreadsheets at this point, don't know how fast we'll get the 'Beast' into any one of those, but maybe there are other solutions to share that information with you. Water portfolio data will work for summary level information within WaDE. Will have that put into an MS Access database by next summer.

Multiple water availability metrics? (Include definitions.)

What is the funding for Phase 2 and when?

What is the schedule for WaterPIE? Currently in the business need scoping phase. Will get from consultant an estimate to build the hub and nodes. Find funding for that in the future. Hoping to roll those into an implementation phase. Version 1 of WQX was about $8 million. Water board project was $1.5M. USGS and EPA web services at waterqualitydata.us.

Required to do an alternative analysis? Yes, but it depends on the size of the project.

Any more questions? Nope… adjourn! ☺